Regression with Gaussian Process Networks

Manel Martínez-Ramón

ECE, UNM

October, 2018

Regression in Hilbert spaces

- So far we have seen the expression of the predictive distribution in linear regression, where

$$p(\bar{f}_*|\mathbf{x}^*, \mathbf{y}, \mathbf{X}) = \mathcal{N}\left(\bar{\mathbf{w}}^\top \mathbf{x}^*, \mathbf{x}^{*\top} \mathbf{A}^{-1} \mathbf{x}^*\right)$$

and

$$\mathbf{A} = \sigma_n^{-2} \mathbf{X}\mathbf{X}^\top + \boldsymbol{\Sigma}_p^{-1}$$

$$\bar{\mathbf{w}} = \left(\mathbf{X}\mathbf{X}^\top + \sigma_n^2 \boldsymbol{\Sigma}_p^{-1}\right)^{-1} \mathbf{X}\mathbf{y}$$

Regression in Hilbert spaces

- Now consider a nonlinear transformation into a Hilbert Space with the form $\boldsymbol{\varphi}(\mathbf{x})$.
- The corresponding expressions for the inverse of the covariance matrix $\mathbf{A}$ and the linear parameters $\mathbf{w}$ are:

$$\mathbf{A} = \sigma_n^{-2} \boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \boldsymbol{\Sigma}_p^{-1}$$

$$\bar{\mathbf{w}} = \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \sigma_n^2 \boldsymbol{\Sigma}_p^{-1} \right)^{-1} \boldsymbol{\Phi}\mathbf{y} \tag{1}$$

where $\boldsymbol{\Phi} = \{\boldsymbol{\varphi}(\mathbf{x}[1]) \cdots \boldsymbol{\varphi}(\mathbf{x}[N])\}$
Note that $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top$ is high dimensional, possibly infinite.

The proofs in Rasmussen et al, 2006 (page 12) are slightly different from the ones presented here.

Regression in Hilbert spaces

- Now we apply the Representer Theorem in the transformation

$$\boldsymbol{\varphi}(x) \rightarrow \boldsymbol{\Sigma}_p^{1/2}\boldsymbol{\varphi}(x)$$

- This is, the parameter vector is a linear combination of the transformed data

$$\mathbf{w} = \boldsymbol{\Sigma}_p^{1/2}\boldsymbol{\Phi}\boldsymbol{\alpha}$$

- Also, the estimator is now

$$\bar{f}_* = \boldsymbol{\varphi}(x^*)^\top \boldsymbol{\Sigma}_p^{1/2}\mathbf{w}$$

Regression in Hilbert spaces

- With the expressions of slide **??**, the estimator can be written as

$$\bar{f}_* = \boldsymbol{\varphi}(x^*)^\top \boldsymbol{\Sigma}_p^{1/2} \boldsymbol{\Sigma}_p^{1/2} \boldsymbol{\Phi} \boldsymbol{\alpha} = \boldsymbol{\varphi}(x^*)^\top \boldsymbol{\Sigma}_p \boldsymbol{\Phi} \boldsymbol{\alpha} \tag{2}$$

- Note that now the dot product between test and training data is $\boldsymbol{\varphi}(x^*)^\top \boldsymbol{\Sigma}_p \boldsymbol{\Phi}$
- Also, note that expression $\boldsymbol{\Sigma}_p \boldsymbol{\Phi} \boldsymbol{\alpha}$ in equation (**??**) plays the role of a new set of parameters, i.e.

$$\bar{f}_* = \boldsymbol{\varphi}(x^*)^\top \boldsymbol{\Sigma}_p \boldsymbol{\Phi} \boldsymbol{\alpha} = \boldsymbol{\varphi}(x^*)^\top \mathbf{w}'$$

Then, we can apply to this new set of parameters the solution of equation (**??**) of slide **??**.

Regression in Hilbert spaces

- This is, for

$$\mathbf{w}' = \boldsymbol{\Sigma}_p \boldsymbol{\Phi} \boldsymbol{\alpha}$$

  we can apply the solution

$$\bar{\mathbf{w}}' = \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \sigma_n^2 \boldsymbol{\Sigma}_p^{-1} \right)^{-1} \boldsymbol{\Phi}\mathbf{y}$$

- and the following equalities hold

$$\boldsymbol{\Sigma}_p \boldsymbol{\Phi} \boldsymbol{\alpha} = \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \sigma_n^2 \boldsymbol{\Sigma}_p^{-1} \right)^{-1} \boldsymbol{\Phi}\mathbf{y}$$

$$\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \sigma_n^2 \boldsymbol{\Sigma}_p^{-1} \right) \boldsymbol{\Sigma}_p \boldsymbol{\Phi} \boldsymbol{\alpha} = \boldsymbol{\Phi}\mathbf{y}$$

Regression in Hilbert spaces

- In order to find $\boldsymbol{\alpha}$ we just need to premultiply it by $\boldsymbol{\Phi}^\top$

$$\boldsymbol{\Phi}^\top \left(\boldsymbol{\Phi}\boldsymbol{\Phi}^\top + \sigma_n^2 \boldsymbol{\Sigma}_p^{-1} \right) \boldsymbol{\Sigma}_p \boldsymbol{\Phi} \boldsymbol{\alpha} = \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \mathbf{y}$$

$$\left(\boldsymbol{\Phi}^\top \boldsymbol{\Phi}\boldsymbol{\Phi}^\top \boldsymbol{\Sigma}_p \boldsymbol{\Phi} + \sigma_n^2 \boldsymbol{\Phi}^\top \boldsymbol{\Sigma}_p^{-1} \boldsymbol{\Sigma}_p \boldsymbol{\Phi} \right) \boldsymbol{\alpha} = \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \mathbf{y}$$

$$\left(\boldsymbol{\Phi}^\top \boldsymbol{\Phi}\boldsymbol{\Phi}^\top \boldsymbol{\Sigma}_p \boldsymbol{\Phi} + \sigma_n^2 \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \right) \boldsymbol{\alpha} = \boldsymbol{\Phi}^\top \boldsymbol{\Phi} \mathbf{y}$$

$$\left(\boldsymbol{\Phi}^\top \boldsymbol{\Sigma}_p \boldsymbol{\Phi} + \sigma_n^2 \mathbf{I} \right) \boldsymbol{\alpha} = \mathbf{y}$$

- Finally

$$\boldsymbol{\alpha} = \left(\boldsymbol{\Phi}^\top \boldsymbol{\Sigma}_p \boldsymbol{\Phi} + \sigma_n^2 \mathbf{I} \right)^{-1} \mathbf{y} = \left(\mathbf{K} + \sigma_n^2 \mathbf{I} \right)^{-1} \mathbf{y}$$

Regression in Hilbert spaces

- In expression $\boldsymbol{\alpha} = \left(\mathbf{K} + \sigma_n^2 \mathbf{I}\right)^{-1} \mathbf{y}$ we assume that

$$\mathbf{K} = \boldsymbol{\Phi}^\top \boldsymbol{\Sigma}_p \boldsymbol{\Phi}$$

  which is a matrix for which entries are

$$k(\mathbf{x}, \mathbf{z}) = \boldsymbol{\varphi}(\mathbf{x})^\top \boldsymbol{\Sigma}_p \boldsymbol{\varphi}(\mathbf{z}) = \boldsymbol{\varphi}(\mathbf{x})^\top \boldsymbol{\Sigma}_p^{1/2} \boldsymbol{\Sigma}_p^{1/2} \boldsymbol{\varphi}(\mathbf{z})$$

  which turns out to be a kernel dot product of vectors in a Hilbert space which have been linearly transformed by matrix $\boldsymbol{\Sigma}_p^{1/2}$. As we proved in the previous chapter, this is a valid dot product.
- The parameter covariance matrix is then hidden in the dot product. This means that the choice if this matrix (which defines the prior on the parameters) is implicit in the choice of the kernel.

Outcomes of this lesson

After this lesson students should be able to

- Derive the solution for the dual parameters $\boldsymbol{\alpha}$ of the Kernel Gaussian process for regression.
- Prove that the kernel dot product between data is

$$k(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}^\top(\mathbf{x}_i)\boldsymbol{\Sigma}_p\boldsymbol{\varphi}(\mathbf{x}_i)$$